Working Paper 2024.1.6.6
- Vol 1, No 6

KHÁM PHÁ TIỀM NĂNG CỦA MÔ HÌNH POLICY GRADIENT PORTFOLIO: CẤU TRÚC HỌC SÂU TĂNG CƯỜNG CHO QUẢN TRỊ DANH MỤC ĐẦU TƯ TẠI THỊ TRƯỜNG CHỨNG KHOÁN VIỆT NAM

Võ Minh Hiếu[1]

Sinh viên K59 CLC Tài chính quốc tế

Trường Đại học Ngoại Thương Cơ sở II tại TP. Hồ Chí Minh

Lê Trung Thành

Giảng viên cơ sở II

Trường Đại học Ngoại Thương Cơ sở II tại TP. Hồ Chí Minh

Tóm tắt

Quản trị danh mục đầu tư tài chính được định nghĩa là việc tạo và duy trì danh mục đầu tư với mục đích đạt được các mục tiêu tài chính cụ thể. Quá trình này bao gồm việc xác định sự kết hợp tối ưu của các loại tài sản khác nhau dựa trên các yếu tố như mục tiêu tài chính của nhà đầu tư, mức độ chấp nhận rủi ro và thời hạn đầu tư. Có nhiều cách tiếp cận khác nhau để quản lý danh mục đầu tư tài chính, từ các phương pháp truyền thống như lý thuyết danh mục đầu tư hiện đại của Harry Markowitz đến các kỹ thuật học máy như học tăng cường. Mặc dù các mô hình học tập tăng cường đã chứng minh sự hiệu quả trong việc quản lý các tài sản có tính thanh khoản cao như tiền điện tử, ngoại hối hoặc cổ phiếu ở các thị trường phát triển, ứng dụng của chúng ở các thị trường chứng khoán cận biên vẫn chưa được nghiên cứu rộng rãi. Trong nghiên cứu này, chúng tôi tinh chỉnh mô hình Policy Gradient Portfolio (PGPortfolio), một mô hình quản lý danh mục đầu tư dựa trên học tập tăng cường hiện đại, cho thị trường chứng khoán Việt Nam. Chúng tôi xem xét hai đặc tính của thị trường chứng khoán Việt Nam, bao gồm thời gian thanh toán T+2 và phí hoa hồng cao. Thử nghiệm của chúng tôi trải dài trong phạm vi dữ liệu 5 năm từ 2018 đến 2023. Kết quả cho thấy mô hình được điều chỉnh có thể hoạt động tốt hơn các quỹ ETF lớn tại Việt Nam, ngay cả khi có chi phí giao dịch cao và hạn chế thanh khoản. Nghiên cứu này góp phần nâng cao hiểu biết về việc áp dụng các kỹ thuật học tăng cường để giải quyết những thách thức cụ thể tại các thị trường chứng khoán cận biên như Việt Nam

[1] Corresponding author, Email: vominhhieu2013346031@ftu.edu.vn

Từ khóa: quản trị danh mục đầu tư, học sâu tăng cường, policy gradient portfolio, thị trường chứng khoán Việt Nam, trí tuệ nhân tạo

EXPLORING THE POTENTIAL OF POLICY GRADIENT PORTFOLIO: A DEEP REINFORCEMENT LEARNING FRAMEWORK FOR PORTFOLIO MANAGEMENT IN THE VIETNAMESE STOCK MARKET

Abstract

Financial portfolio management is defined as the creation and maintenance of an investment portfolio with the aim of achieving specific financial objectives. This process involves determining the optimal mix of asset classes based on factors like an investor's financial goals, risk tolerance, and investment horizon. Various approaches exist for financial portfolio management, ranging from traditional methods like Harry Markowitz's Modern Portfolio Theory to machine learning techniques such as reinforcement learning. While existing reinforcement learning-based models have proven effective for managing highly liquid assets like cryptocurrencies, forex, or stocks in developed markets, their application in frontier stock markets has not been extensively explored. In this study, we adapt and fine-tune the Policy Gradient Portfolio (PGPortfolio), a state-of-the-art reinforcement learning-based portfolio management model, for the stock market in Vietnam. We aim to consider two unique characteristics of the Vietnam stock market, including a T+2 settlement period and high commission fees. Our experiment spans a 5-year data range from 2018 to 2023. The results indicate that the adapted model can outperform certain elite passive fund benchmarks, even in the presence of high transaction costs and liquidity restrictions. This study contributes to the understanding of applying reinforcement learning techniques to address specific challenges in frontier stock markets like Vietnam

Keywords: portfolio management, deep reinforcement learning, policy gradient portfolio, Vietnam stock market, artificial intelligence

1. Introduction

Portfolio management is concerned with effectively distributing resources among various asset classes to achieve optimal returns while managing risks. The theoretical foundations of portfolio management have been established for a long time. The concept was initially introduced by Markowitz (1952) through the mean-variance portfolio optimization framework. This framework enables investors to allocate their wealth across different assets in a manner that balances the risk-reward trade-off according to their risk tolerance. Over time, various authors have further developed this approach, and related methods continue to be widely utilized in both industry and academia. Some studies include the Capital Asset Pricing Model (CAPM), Black Litterman Model, and Fama-French Three-Factor Model (Sharpe, 1964; Black & Litterman, 1992; Fama & French, 1993).

Although traditional strategies show effectiveness under specific circumstances, they encounter challenges in others, primarily due to their limited adaptability to market changes. This limitation stems from their reliance on human cognitive processes and responsiveness to human expectations. Investors have demonstrated irrational behavior, frequently leading to wrong investment decisions. To address the limitations of human decision-making, numerous investment strategies grounded in machine learning have been introduced. Classical machine learning algorithms such as LSTM and CNN have been explored for predicting stock prices or identifying market patterns, demonstrating favorable outcomes (Mukherjee et al., 2021; Nelson et al., 2017). However, a significant portion of these approaches focuses solely on price forecasting without providing a comprehensive portfolio management strategy. Additionally, there are concerns regarding the accuracy and robustness of these models in the face of dynamic market conditions.

As a response, reinforcement learning emerges as a viable alternative to classical deep learning portfolio management models. Numerous frameworks in deep reinforcement learning have been proposed to develop effective portfolio management strategies. These frameworks have shown success in handling highly liquid assets like cryptocurrencies, forex, or US stocks (Ye et al., 2020; Wang et al., 2021; Usha et al., 2019). However, there is a lack of applications in frontier stock markets. Frontier stock markets present distinct characteristics that may pose challenges to existing reinforcement learning frameworks. These challenges include liquidity restrictions, high risk levels, high transaction costs and limited data availability. Examining these challenges and addressing the research gap in applying deep reinforcement learning to frontier stock markets could be a valuable area for further investigation by researchers.

The Vietnam stock market is listed as one of the frontier markets by MSCI and FTSE Russell. Established in 2000, the Vietnam stock market is still a young stock market but has experienced significant growth and development over the years. The VNINDEX, or Vietnam Ho Chi Minh Stock Index, is the primary stock market index that reflects the performance of the Ho Chi Minh City Stock Exchange (HOSE) in Vietnam. Vietnam has established a self-imposed deadline of 2025 to upgrade its stock market to an emerging stock market. However, there is a risk of missing this deadline due to internal conflicts among state institutions regarding pivotal reforms, such as those related to settlements and foreign ownership of companies. In Vietnam, the stock trading system enforces a T+2 settlement cycle, meaning the transfer of securities and funds occurs two business days after the trade date. This restriction makes the market illiquid and then poses challenges to frequent trading strategies. In addition, concerns also exist about the quality of data, which affects the development of machine-learning-based investment approaches.

There is previous research to study the application of reinforcement learning in portfolio management for the Vietnam stock market. For example, Ngo et al. (2009) conducted a comparative analysis between reinforcement learning and classical approaches in portfolio management within the Vietnam stock market. Another study by Nguyen et al. (2009), introduced a novel reinforcement learning framework incorporating customized technical indicators tailored for the specific characteristics of the Vietnam stock market. Although these research findings exhibit positivity, a research gap remains unaddressed—there exists no explicit examination of policy gradient approaches concerning frequent trading of underlying securities. Frequent trading of underlying securities may subject investors to higher transaction costs compared to derivative securities. Also, it has to deal with the T+2 settlement restriction.

This paper aims to assess the effectiveness and profitability of the Policy Gradient Portfolio (PGPortfolio) within the context of the Vietnam stock market. The Policy Gradient Portfolio is a deep reinforcement learning framework introduced by Jiang et al. (2017), recognized as a leading framework for portfolio management. Modifications were made to adapt the original framework to the specifics of the Vietnam stock market. Backtesting was conducted under two distinct market trends (bullish and bearish) to evaluate the model's robustness. The evaluation incorporates four financial metrics: Sharpe ratio, Information ratio, Max drawdown, and Cumulative portfolio value.

The model's performance is then compared against the VN-INDEX benchmark and five different passive ETF funds in Vietnam. The experimental results demonstrated that PGPortfolio not only exceeded market performance but also surpassed similar benchmarks.

2. Research background

2.1. Reinforcement learning

Reinforcement learning (RL) is a framework for sequential decision-making, wherein agents learn to take actions within an environment with the objective of maximizing cumulative rewards (Prince, 2023). This learning process is unsupervised, meaning there is no explicit instruction provided by humans to the agent. Instead, agents learn from feedback received from the environment as a consequence of specific actions taken. For example, in finance, an RL algorithm might assist a trader (the agent) to engage in buying or selling assets (the actions) on a financial market (the environment) to maximize profit (the reward). When the trader executes a trade based on its market trend analysis and incurs losses, it considers it a failure, revisits the trade, learns from the failure, and adjusts its strategy to avoid similar trades in the future.

2.2. Reinforcement learning in portfolio management

Previous works show that reinforcement learning can be effectively applied in financial portfolio management. In the context of portfolio management, fundamental components of the reinforcement learning framework consist of a virtual trader as the agent, the financial market as the environment, the representation of the market as the state, and the trading behavior as the action. The environment can be various financial markets, spanning stock markets, cryptocurrencies, or forex (Ye et al., 2020; Wang et al., 2021; Usha et al., 2019). The state includes a variety of indicators that capture the market trend, encompassing financial metrics, consolidated historical prices, existing asset allocations, or correlations among assets (Jiang et al., 2017; Lillicrap et al., 2015; Durall, 2022). The action in these studies also exhibits variability, ranging from discrete actions such as buy, sell, or hold, to continuous actions that define the allocation of different asset classes within the financial portfolio (Wang et al., 2021; Zhang et al., 2020).

In terms of reinforcement learning algorithms, there are also different algorithms that prove effective for portfolio management. The most popular learning algorithm is Deep Deterministic Policy Gradient (DDPG) (Jiang et al., 2017). Some works made use of more recent algorithms that also improve upon Policy Gradients, such as Proximal Policy Optimization (PPO), Advantage Actor-Critic (A2C), and Soft Actor-Critic (SAC) (Durall, 2022). Despite showing high potential to outperform DDPG, recent algorithms are too complicated to be applied in the frontier stock market like Vietnam where the limited data availability can undermine their potential.

Policy Gradient Portfolio (PGPortfolio) leverages on the Deep Deterministic Policy Gradient algorithm. It simplifies the algorithm when using only one target network instead of an actor-critic architecture. It defines the action as the allocation of assets within the portfolio at each rebalance time. The state is represented by historical price data at each time frame, which undergoes preprocessing before being input into the model. The deep neural networks can include CNN, LSTM, or RNN. The original paper employed an ensemble approach which constructed a system of three identical reinforcement learning frameworks. The initial experiment focused on a portfolio comprising 12 cryptocurrencies, but the framework's applicability extends to various financial markets. In this study, to simplify, we employ only CNN to assess its efficacy within the Vietnam stock market.

3. Research method

3.1. Overview of policy gradient portfolio

The core of the policy gradient portfolio centers around three elements: market state representation, neural policy network and reinforcement learning algorithms. The original framework allowed the incorporation of transaction costs and required two initial assumptions, including:

a) Assuming high liquidity for all market assets, each trade can be executed immediately at the last price when an order is placed.

b) It is assumed that the capital invested by the virtual trader is so insignificant that it has no influence on the market.

For the test case of Vietnam, we remain two above assumptions and incorporate the transaction cost ranging from 0% to 0.25%.

3.1.1. Market state representation

The framework uses historical price data to represent the market state. The historical price data is input into the neural network, generating the output in the form of a portfolio vector. Specifically, the input to the neural policy network at the end of period t is represented as a tensor with dimensions (b, f, n, l), where b denotes the batch size, f indicates the number of features, n represents the count of non-cash financial assets included in the portfolio, and l signifies the number of selected lags for historical data. The closing prices, lowest prices, and highest prices for each trading session are selected as features. A notable point is that the neural networks do not receive the absolute price values directly as input. Instead, these values are normalized using the latest closing prices, as indicated by the following equations.

$$X_t = [V_t^{(close)}\ V_t^{(high)}\ V_t^{(low)}]\ (1)$$

$$V_t^{(close)} = [\frac{v_{t-1+1}^{(close)}}{v_t^{(close)}}\ \frac{v_{t-n+2}^{(close)}}{v_t^{(close)}} \cdots\ \frac{v_{t-1}^{(close)}}{v_t^{(close)}}\ \frac{v_t^{(close)}}{v_t^{(close)}}]\ (2)$$

$$V_t^{(high)} = [\frac{v_{t-1+1}^{(high)}}{v_t^{(close)}}\ \frac{v_{t-n+2}^{(high)}}{v_t^{(close)}} \cdots\ \frac{v_{t-1}^{(high)}}{v_t^{(close)}}\ \frac{v_t^{(high)}}{v_t^{(close)}}]\ (3)$$

$$V_t^{(low)} = [\frac{v_{t-1+1}^{(low)}}{v_t^{(close)}}\ \frac{v_{t-n+2}^{(low)}}{v_t^{(close)}} \cdots\ \frac{v_{t-1}^{(low)}}{v_t^{(close)}}\ \frac{v_t^{(low)}}{v_t^{(close)}}]\ (4)$$

3.1.2. Neural policy network

At the end of the period $t-1$, the virtual portfolio manager uses the neural policy network π to generate a portfolio vector w_t, utilizing the price tensor X_{t-1} and the previous portfolio vector w_{t-1}. The portfolio vector w_t signifies the allocation weight of financial assets in the portfolio at the beginning of period t. Mathematically, $w_t = \pi(X_{t-1}, w_{t-1})$. As mentioned, the original framework ensembled three types of neural networks to get the final result. However, in the specific context of the Vietnam stock market, we opt to solely utilize the CNN topology. Figure (1) visualizes the architecture of the neural policy network, in which the input is the price tensor X_{t-1} along with the previous allocation w_{t-1}, and the output is the allocation for the upcoming period w_t.

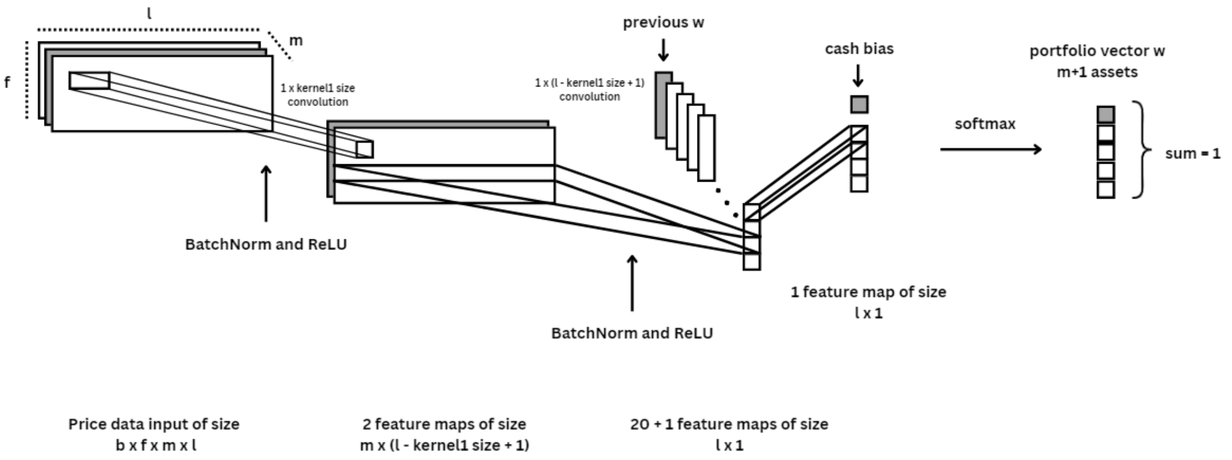


Figure 1: PGPortfolio architecture with a CNN-based neural policy network.

Source: Jiang et al. (2017)

The convolutional neural network (CNN) architecture consists of 3 weight layers, where the first dimension of the local small receptive field for all convolutional layers is 1. The first shape of kernels is set to 1 because the convolutional layers are only considering temporal information of each asset. In CNNs, the movement of the kernel refers to how the filter or kernel slides over the input data during the convolution operation. Because our input tensor has the shape of (f, m, l) (excluding the batch size) and a column vector represents the feature value for each time step, if the first dimension of the kernel is set to 1, it means that the kernel has a width of 1 in terms of the temporal dimensions and the convolution operation is applied independently along each temporal position of the input data.

Each row within the entire network is designated to a specific asset and is tasked with calculating a voting score after 3 convolutional operations. Such voting scores are then submitted to a softmax layer in order to outcome the portfolio weight vector for the next period. Indicated by equation (5), a softmax function is applied to normalize the values into the range of (0,1). Notice that before the softmax function is applied, a cash bias is concatenated to the feature map of dimensions (m, l) to achieve a complete portfolio. Cash bias, therefore, is defined as one of the weights of the neural network.

$$s(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \quad (5)$$

3.1.3. Reinforcement learning algorithms

As described above, the reinforcement learning algorithm is to find out the optimal policy guiding actions based on the state to maximize rewards. Given the portfolio weight vector computed by the neural policy network, the reward is calculated as the logarithmic rate of return of the portfolio for the next period. The period t is defined as the time between the opening price at time t and the opening price at time $t + 1$. It is also equivalent to the time between the closing price at time $t - 1$ and the closing price at time t. The calculation of the return for the period t is identified by equation (6).

$$R_t = \frac{P_{t+1} - P_t}{P_t}, t \geq 1 \text{ (6)}$$

Where P_t is the portfolio value at the beginning of period t ($t \geq 1$) (at the time $t - 1$), P_{t+1} is the portfolio value at the beginning of period $t + 1$ (at the time t) and $R_t$is the rate of return for the period t. The value of the portfolio at time t is the sum of the values of all financial assets within the portfolio at that time. The calculation of the value of each asset at time t, denoted as $p_{i,t}$, relies on its value at time $t - 1$ and the associated rate of return for period $t - 1$. The value of each asset at time $t - 1$ is determined by multiplying the value of the entire portfolio at time $t - 1$ by its corresponding allocation weight at that time, $w_{i,t-1}$. Meanwhile, the return rate for each asset in period $t - 1$ is precisely expressed by the relative price vector quantity y_{t-1} for the $(t - 1)$-th trading period, calculated as the value of dividing the closing price of period $t - 1$ (at the time t), denoted as v_t by the closing price of period $t - 2$ (at the time $t - 1$), denoted as v_{t-1}:

$$P_{t+1} = \sum_i^m p_{i,t} \text{ (7)}$$

$$p_{i,t} = p_{i,t-1}\, r_{i,t-1} \text{ (8)}$$

$$p_{i,t-1} = P_t w_{i,t-1} \text{ (9)}$$

$$r_{i,t-1} = \frac{v_{i,t}}{v_{i,t-1}} \text{ (10)}$$

We can rewrite the calculation of the portfolio value at the beginning of period t and the computation of portfolio return of the period t as follows:

$$P_{t+1} = P_t \times \sum_i^m w_{i,t-1} \frac{v_{i,t}}{v_{i,t-1}} \text{ (11)}$$

$$R_t = \sum_i^m w_{i,t-1} \frac{v_{i,t}}{v_{i,t-1}} - 1 \text{ (12)}$$

However, such formulas ignore the transaction cost. In the real world, each trade incurs a commission fee ranging from 0% to 0.25% of the transaction value. At the end of the period $t - 1$, the virtual portfolio manager must adjust the portfolio vector from $w'_{t-1}$to w_t where w'_{t-1}represents the portfolio weight vector at the end of period $t - 1$ and $w_t$describes the portfolio weight vector at the beginning of period t. Following the payment of all commissions, this reallocation results in a reduction of the portfolio value by the factor μ_t, representing the transaction remainder factor. Figure (2) shows how the allocation process works. The adjusted portfolio value at the beginning of period t is:

$$P_t = \mu_t P'_{t-1} \text{ (13)}$$

The return rate for the period t now is rewritten as follows:

$$R_t = \mu_t \times \sum_i^m w_{i,t-1} \frac{v_{i,t}}{v_{i,t-1}} - 1 \text{ (14)}$$

The cumulative rate of return at the end is:

$$CR_{end} = \prod_{t=1}^{t_{end}} R_t = \prod_{t=1}^{t_{end}} \left(\mu_t \times \sum_i^m w_{i,t-1} \frac{v_{i,t}}{v_{i,t-1}} \right) \text{ (15)}$$

And the portfolio value at the end is:

$$P_{end} = P_1 CR_{end} = P_1 \prod_{t=1}^{t_{end}} \left(\mu_t \times \sum_i^m w_{i,t-1} \frac{v_{i,t}}{v_{i,t-1}} \right) \text{ (16)}$$

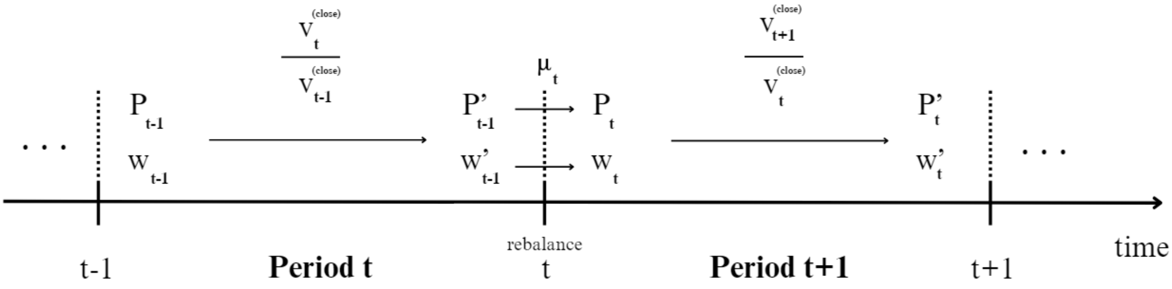


Figure 2: The allocation timeline with the effect of transaction remainder factor μ_t.

Source: Jiang et al. (2017)

Equation (16) also represents the reward received after each episode. The objective of reinforcement learning algorithms is to learn to maximize the reward. Therefore, the objective function is the logarithmic cumulative rate of return at the end of each episode. For the application of the gradient-based optimization method, Equation (15) is reformulated as equation (17), signifying the loss function of the neural policy network.

$$L = -ln(CR_{end}) = -ln(\prod_{t=1}^{t_{end}} (\mu_t \times \sum_i^m w_{i,t-1} \frac{v_{i,t}}{v_{i,t-1}}) \quad (17)$$

3.2. A modification tailored for the Vietnam stock market

A modification is made to the loss function. While the original framework employs the logarithmic cumulative return of the portfolio itself, we employ the difference between the cumulative Sharpe ratio of the portfolio and that of the benchmark VNINDEX. The difference between the portfolio and the benchmark makes more sense because a good portfolio manager is considered to win the market. The Sharpe ratio which was proposed by Sharpe (1994) is used as a primary measure to evaluate the risk-adjusted return of the portfolio. Incorporating it into the loss function aims to make the virtual trader more aware of the risk element in its decision-making process. Let R_p and R_f denote the rate of return of the portfolio and the risk-free asset, respectively. Equation (18) and (19) show the calculation of the Sharpe ratio and how the loss function is modified.

$$ShR = \frac{E(R_p - R_f)}{\sqrt{var(R_p)}} \quad (18)$$

$$L = -(ShR_p - ShR_{VNI}) \quad (19)$$

3.3. Experiment setups

3.3.1. Data selection

In the experiment for the Vietnam stock market, the portfolio consists of the 14 companies chosen randomly from the VN30 list on the HOSE (in 07/2021). Along with cash, it results in a total size of the portfolio denoted as m + 1, which is equal to 15. The stocks from the VN30 list are chosen with the goal of meeting two primary assumptions. Firstly, these stocks are among the most traded in Vietnam, ensuring a high level of liquidity, allowing each trade to be promptly executed at the last price when an order is placed (assumption 1). Additionally, their substantial market capitalization ensures that any action taken by the virtual trader is of such insignificance that it has no impact on the overall market (assumption 2). Appendix 1 lists all of the selected stocks for the experiment.

In this research, we get historical price data from the Investing's database. The dataset, from 24/02/2015 to 19/12/2023, consists of 2174 observations of daily OHLC data. To facilitate our analysis, we divide the dataset into distinct training and testing sets. The training set includes 1933 observations, from 24/02/2015 to 18/11/2022. The testing set includes 241 observations, ranging from 03/01/2023 to 19/12/2023. There is a gap between the training set and the testing one. This is because the first testing step starts on 03/01/2023 and it requires 30 lag values to input to the neural policy network. To check the robustness of the virtual trader, we do a back-test experiment twice. The first attempt is to cover the whole 241 time steps of the testing dataset, equivalent to the entire year. The second experiment is only carried out to cover a second half of the year 2023. While the entire year 2023 witnessed a growth of VNINDEX by more than 5%, the latter half experienced a sharp decline. This choice of back-test experiments is motivated by the observation of distinct market trends during 2023, and we aim to evaluate the framework's performance under different conditions.

3.3.2. Evaluation metrics

A set of standard financial metrics is used to assess the performance of PGPortfolio. It includes Sharpe ratio, Max drawdown, Cumulative portfolio value, and Information ratio. As mentioned above, the Sharpe ratio serves as the principal metric for assessing the risk-adjusted return of the portfolio. This metric measures the excess return per unit of risk, providing an objective assessment of the portfolio's efficiency in generating returns relative to its associated risk.

Max drawdown evaluates the largest loss experienced by the portfolio during a specified period (Magdon-Ismail & Atiya, 2004). This metric is also particularly relevant for assessing the risk tolerance and resilience of the portfolio strategy, with a lower max drawdown indicating a more stable and robust portfolio.

$$MDD = \frac{Trough\ value - Peak\ value}{Peak\ value} \quad (20)$$

The Cumulative Portfolio Value represents the cumulative performance of the portfolio over the evaluation period. This metric provides a comprehensive overview of the portfolio's growth and is essential for understanding its overall profitability. Equation (16) describes the calculation of the cumulative portfolio value at the end of the period. In this experiment, the initial investment is set to $10,000 and the higher cumulative portfolio value at the end indicates the better performance.

Finally, the Information ratio is incorporated to measure the portfolio manager's ability to generate excess returns relative to a benchmark index. This ratio assesses the manager's skill in outperforming the market and offers insights into the effectiveness of the active management strategy (Goodwin, 1998). It provides a comprehensive evaluation to our framework because our objective function is built on the excess risk-adjusted returns of the portfolio relative to VNINDEX. Let σ_{p-VNI} denote the tracking error, which is the standard deviation of the difference between the portfolio and VNINDEX returns. Equation (21) formulates the computation of the Information ratio.

$$InR = \frac{R_p - R_{VNI}}{\sigma_{p-VNI}} \quad (21)$$

4. Results and discussion

4.1. Results

The back-testing experiments are run on two different time ranges with five escalated transaction costs, including 0.00%, 0.10%, 0.15%, 0.20%, and 0.25%. The results of the PGPortfolio and other benchmark portfolios are presented in Table 1 and 2. Table 1 illustrates the performance of the portfolio throughout the entire year of 2023, while Table 2 displays the results during a bearish trend in the second half of the year.

4.1.1. Evaluation for the entire year 2023

Table 1: The results of the experiment for the entire year 2023

	Name	**ShR**	**MDD**	**InR**	**CumVal**
1	VNINDEX	0.0246	-0.1745		1.0502
2	FUESSV50	0.0457	-0.2142	**0.0335***	**1.2048***
3	FUESSV30	0.0245	-0.1612	0.0082	1.0621
4	FUEMAV30	0.0243	-0.1739	0.0050	1.0527
5	FUEVFVND	0.0340	-0.1674	0.0195	1.0863
6	FUEVN100	0.0408	-0.1583	0.0272	1.1016
7	PGPortfolio - 0.00%	**0.0860*****	**-0.0981*****	**0.0544*****	**1.3168*****
8	PGPortfolio - 0.10%	**0.0685****	**-0.1065****	**0.0402****	**1.2375****
9	PGPortfolio - 0.15%	**0.0596***	**-0.1108***	0.0331	1.1996
10	PGPortfolio - 0.20%	0.0507	-0.1150	0.0260	1.1628
11	PGPortfolio - 0.25%	0.0417	-0.1191	0.0188	1.1271

(***): 1st highest; (**): 2nd highest; (*): 3nd highest

ShR: Sharpe ratio; **MDD**: Max drawdown; **InR**: Information ratio; **CumVal**: Cumulative value

Source: Compiled by the author

In general, the PGPortfolio demonstrates superior performance compared to other benchmarks, achieving high returns with lower associated risks. Despite the impact of transaction costs on performance, the PGPortfolio consistently outperforms the market, even at the highest transaction cost of 0.25%. The PGPortfolio for high commission fees outperforms all benchmarks, except for the FUESSV50, showcasing significant dominance across various financial metrics. It achieves double-digit returns, while maintaining a relatively low max drawdown ranging from -10% to -12%. In contrast, other benchmarks exhibit less favorable returns, with returns of less than 10% and higher max drawdowns exceeding -15%. The Sharpe ratio of the PGPortfolio, which is integrated into the loss function, is consistently at least twice that of the VNINDEX benchmark. Even in comparison to the FUESSV50 benchmark, the PGPortfolio still proves more effective. Although the FUESSV50 achieves a yearly return of over 20%, it is exposed to high risks indicated by a max drawdown of -21%.

4.1.2. Evaluation for the second half of the year 2023

Table 2: The results of the experiment for the second half of the year 2023

	Name	ShR	MDD	InR	CumVal
1	VNINDEX	-0.0378	-0.1745		0.9407
2	FUESSV50	-0.0115	-0.2142	0.0114	0.9534
3	FUESSV30	-0.0347	-0.1612	0.0006	0.9401
4	FUEMAV30	-0.0293	-0.1739	0.0221	0.9496
5	FUEVFVND	0.0035	-0.1674	**0.0570****	0.9936
6	FUEVN100	-0.0071	-0.1583	**0.0564***	0.9833
7	PGPortfolio - 0.00%	**0.0476*****	**-0.0981*****	**0.0607*****	**1.0766*****
8	PGPortfolio - 0.10%	**0.0287****	**-0.1065****	0.0455	**1.0390****
9	PGPortfolio - 0.15%	**0.0192***	**-0.1108***	0.0379	**1.0207***
10	PGPortfolio - 0.20%	0.0096	-0.1150	0.0303	1.0026
11	PGPortfolio - 0.25%	0.0001	-0.1191	0.0226	0.9850

(***): 1st highest; (**): 2nd highest; (*): 3nd highest

ShR: Sharpe ratio; **MDD**: Max drawdown; **InR**: Information ratio; **CumVal**: Cumulative value

Source: Compiled by the author

The latter part of 2023 experienced a sharp downturn from August to October, and it just started to gradually recover towards the year's end. In the absence of risk considerations, a portfolio manager may gain significantly during a bullish trend but incur substantial losses in a bearish one. Therefore, it is crucial to assess the PGPortfolio's performance during this bearish trend. The objective function, incorporating the Sharpe ratio, is expected to assist the virtual trader in reaching a balance between risk and return in its decision-making process. Table 2 shows that the PGPortfolio meets such expectations. Despite the negative growth experienced by all comparative benchmarks during the downturn, the PGPortfolio still managed to secure positive returns, except when incorporating a transaction cost of 0.25%. As expected, the benchmark FUESSV50, with high risk exposure, no longer dominated the other benchmarks. Its cumulative value decreased to 0.95, and the Sharpe ratio reached only -0.01. The most robust benchmark during this period was FUEVFVND, exhibiting a positive Sharpe ratio and a cumulative value of 0.99. However, this performance is still less impressive compared to that of the PGPortfolio. The PGPortfolio achieved positive returns with lower associated risks during the bearish trend. At transaction costs lower than 0.25%, it outperformed the best benchmark, FUEVFVND, across all financial metrics (except for the Information ratio). At a transaction cost of 0.25%, FUEVFVND surpassed the PGPortfolio in two metrics—Sharpe ratio and Information ratio; but it earned lower cumulative returns and a higher max drawdown compared to the PGPortfolio.

4.2. Discussions

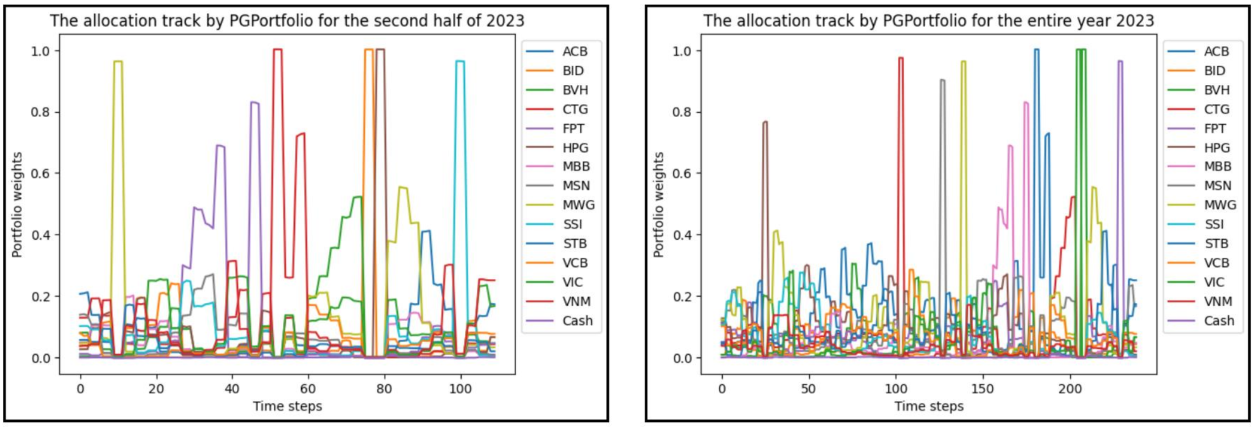
The superiority of the PGPortfolio over comparative benchmarks in this study provides a crucial indication of the heightened capability offered by deep reinforcement learning framework compared to traditional portfolio management methodologies. However, certain limitations still existed, prompting the need for additional investigation into these issues. Real portfolio managers should consider (1) the application of PGPortfolio in the real world, especially when Vietnam still places some restrictions on the stock market. Future research efforts should also concentrate on further refining deep learning models to deal with (2) asset overweight.

4.2.1. Real-world application

The practical implementation of PGPortfolio in Vietnam is contingent upon the aforementioned two initial assumptions. This implies that the PGPortfolio framework may not be suitable for utilization by portfolio managers in large investment firms, such as Dragon Capital or VinaCapital. Instead, individual investors appear to be more fitting candidates for effectively deploying the PGPortfolio framework. It is essential to note that individual investors face elevated transaction costs, and experimental results indicate that the performance of PGPortfolio is adversely impacted by heightened transaction fees. Therefore, any real-world application of the framework should consider the applicable transaction cost policy and carefully select stocks that exhibit sufficient liquidity to meet the initial assumptions.

4.2.2. Asset overweight

Figure (3) shows the allocation track of 14 stocks in the portfolio during experiments. The agent consistently allocates a significant portion of funds to specific stocks across various timeframes. Overweighting in certain assets may elevate the portfolio's exposure to higher volatility, particularly if these assets exhibit greater volatility than the market average. This increased volatility can result in more substantial price fluctuations within the portfolio. Therefore, it is imperative to manage the allocation of each stock within a secure range to mitigate the risk of the portfolio being exposed to excessive volatility. However, our virtual trader is not equipped to execute such allocation control. Instead, its primary objective is to maximize excess risk-adjusted returns. If a specific stock demonstrates notably high excess risk-adjusted returns, it is likely that the agent will concentrate its allocation towards such stocks. To tackle this problem, further research can incorporate a regularization term into the loss function, imposing a penalty on significant weights assigned to particular specific classes.



(a)

(b)

Figure 3: The portfolio weights allocated by PGPortfolio during two experiments given the transaction cost of 0.15%.

Source: Compiled by the author

5. Conclusion

In this study, we evaluate the efficacy of PGPortfolio, a powerful deep reinforcement learning framework, in the context of the Vietnam stock market. A simplified iteration of the original framework is applied to manage a financial portfolio consisting of the top 14 traded stocks in Vietnam. The virtual portfolio manager seeks to maximize risk-adjusted returns and outperform comparable benchmarks by introducing the excess Sharpe ratio into the loss function. Experimental scenarios are conducted to assess PGPortfolio's performance in both bullish and bearish market trends in 2023. These experiments involve simulations where the virtual agent must navigate settlement restrictions and deal with high commission fees.

Five evaluation metrics were employed to appraise the performance of the portfolio. The experimental outcomes revealed that PGPortfolio not only surpassed the market but also outperformed comparable benchmarks. Throughout the entirety of 2023, where the market index witnessed a modest 5% growth and other ETF fund indices achieved growth ranging from 5% to 20%, PGPortfolio achieved a remarkable return rate of 31% when ignoring transaction costs. Accounting for transaction costs did impact its performance, yet it continued to demonstrate effectiveness compared to alternative benchmarks. Notably, PGPortfolio achieved these impressive returns with significantly lower associated risks compared to the market and benchmarks. During the bearish trend in the latter half of 2023, PGPortfolio stood as the only portfolio exhibiting positive growth, while others observed declines in their portfolio values. This suggests a superior risk consideration, as evidenced by its smaller maximum drawdown. Furthermore, its Sharpe ratio, and Information ratio provide additional evidence of the effective trading strategies employed by PGPortfolio. The significance of comprehending frontier markets is growing, and the utilization of advanced technologies like reinforcement learning unquestionably plays an increasingly pivotal role in undertaking such endeavors.

REFERENCES

Markowitz, H. M. (1952). “Portfolio selection”, *Journal of Finance,* Vol. 7 No. 1, pp. 77-91.

Sharpe, W. F. (1964). “Capital asset prices: A theory of market equilibrium under conditions of risk”, *Journal of Finance*, Vol. 19 No. 3, pp. 425-442.

Black, F. & Litterman, R. (1992). “Global portfolio optimization”, *Financial Analysts Journal*, Vol. 48 No. 5, pp. 28-43.

Fama, E. F. & French, K. R. (1993). “Common risk factors in the returns on stocks and bonds”, *Journal of Financial Economics,* Vol. 33 No. 1, pp. 3-56.

Mukherjee, Somenath, Bikash S., Nairita S., Debajyoti R. & Soumil D. (2021). “Stock market prediction using deep learning algorithms”, *CAAI Transactions on Intelligence Technology*, Vol. 8 No. 1, pp. 82-94.

Nelson, D. M., Pereira, A. C. & De Oliveira, R. A. (2017). “Stock market's price movement prediction with LSTM neural networks”, *2017 International joint conference on neural networks (IJCNN),* pp. 1419-1426.

Ye, Y., Pei, H., Wang, B., Chen, P. Y., Zhu, Y., Xiao, J. & Li, B. (2020). “Reinforcement-learning based portfolio management with augmented asset movement prediction states”, *Proceedings of the AAAI Conference on Artificial Intelligence,* Vol. 34 No. 01, pp. 1112-1119.

Wang, Z., Huang, B., Tu, S., Zhang, K & Xu, L. (2021). “DeepTrader: a deep reinforcement learning approach for risk-return balanced portfolio management with market conditions Embedding”, *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35 No. 1, pp. 643-650.

Usha, B. A., Manjunath, T. N. & Mudunuri, T. (2019). “Commodity and Forex trade automation using deep reinforcement learning, *2019 1st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE)*, pp. 27 - 31

Jiang, Z., Xu, D. & Liang, J. (2017). “A deep reinforcement learning framework for the financial portfolio management problem”.

Prince, S. J. D. (2023). “Reinforcement learning”, *Understanding Deep Learning*, pp. 373-378

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., ... & Wierstra, D. (2015). Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971.

Durall, R. (2022). “Asset Allocation: From Markowitz to Deep Reinforcement Learning”.

Zhang, Z., Zohren, S. & Stephen, R. (2020). “Deep reinforcement learning for trading”, *The Journal of Financial Data Science.*

Ngo, V. M., Nguyen, H. H. & Van Nguyen, P. (2023). “Does reinforcement learning outperform deep learning and traditional portfolio optimization models in frontier and developed financial markets?”, *Research in International Business and Finance*, Vol. 65, p. 101936.

Nguyen, H. T. N., Mac, B. N. N., Tran, A. D., Nguyen, N. T. & Pham, D. T. (2022). Deep Reinforcement Learning Approach Using Customized Technical Indicators for A Pre-emerging Market: A Case Study of Vietnamese Stock Market”, *2022 RIVF International Conference on Computing and Communication Technologies (RIVF),* pp. 737-742.

Sharpe, W. F. (1994). "The Sharpe ratio", *Journal of Portfolio Management*, Vol. 21 No. 1, pp. 49-58.

Magdon-Ismail, M. & Atiya, A. F. (2004). “Maximum drawdown”, *Risk Magazine*, Vol. 17 No. 10, pp. 99-102.

Goodwin, T. H. (1998). “The information ratio”, *Financial Analysts Journal*, Vol. 54 No. 4, pp. 34-43.

APPENDIX

Appendix 1. List of selected Vietnam stocks in the experimented portfolio

Index	Ticker	Description
1	ACB	Asia Commercial Joint Stock Bank
2	BID	Joint Stock Commercial Bank for Investment and Development of Vietnam (BIDV)
3	BVH	BaoViet Holding
4	CTG	Vietnam Joint Stock Commercial Bank for Industry and Trade (VietinBank)
5	FPT	FPT Corporation
6	HPG	Hoa Phat Group
7	MBB	Military Commercial Joint Stock Bank
8	MSN	Masan Group
9	MWG	Mobile World Investment Corporation (The Gioi Di Dong)
10	SSI	SSI Securities Corporation

11	STB	Sai Gon Thuong Tin Joint Stock Commercial Bank (Sacombank)
12	VCB	Joint Stock Commercial Bank for Foreign Trade of Vietnam (Vietcombank)
13	VIC	Vingroup Joint Stock Company
14	VNM	Vinamilk

Source: Compiled by the author